

COMPOUND WORD BREAKER AND SPELL CHECKER

ABSTRACT OF THE DISCLOSURE

A method of determining the component words of a
5 compound word is disclosed. The method identifies the
component words, by comparing the word with a list of
words found in a lexicon. If the word is not found in
the lexicon the method proceeds to analyze the word
on a character-by-character basis. After each
10 character the method identifies any potential matches
to the selected characters in the lexicon. If a match
is found, it is added to a hypothesis trace in a
lattice. Next, the method checks to see whether the
remaining characters form a valid entry in the
15 lexicon, and whether the entry is an allowed to be a
final segment: All encountered component words are
entered into the lattice, thus creating possibly more
than one hypothesis path. Some paths may be rendered
invalid, if they don't contain the required "seg1"
20 annotation for non-final segments or had encountered
an "anti-seg" bit for presumed final segment. The
output can be ranked if more than one valid
segmentation is found. The method can also correct
spelling errors due to incorrect compounding.